

Simple perfect samplers using monotone birth-and-death processes*

Hiroyuki Masuyama[†]

Department of Systems Science, Graduate School of Informatics, Kyoto University
Kyoto 606-8501, Japan

Abstract

This paper proposes simple perfect samplers using monotone birth-and-death processes (BD-processes), which draw samples from an arbitrary finite discrete target distribution. We first construct a monotone BD-process whose stationary distribution is equal to the target distribution. We then derive upper bounds for the expected coalescence time of the copies of the monotone BD-process. We also establish upper bounds for the expected values and tail probabilities of the running times of two perfect samplers, which are *Doubling CFTP* and *Read-once CFTP* using our monotone BD-process. The latter sampler can draw samples *exactly* from unnormalized target distributions with little memory consumption.

Keywords: Perfect sampling; Coupling from the past (CFTP); Monotone Markov chain; Birth-and-death process (BD-process); Doubling CFTP; Read-once CFTP

Mathematics Subject Classification: Primary 65C05, 65C10; Secondary 60J10, 60J22.

1 Introduction

Perfect sampling algorithms are based on “Coupling From The Past (CFTP)”, proposed by Propp and Wilson (1996). CFTP is a powerful technique that enables us to perform *perfect sampling* from the target distribution, i.e., to generate, in a finite time, samples that *perfectly* follow the target distribution. Basically, CFTP is time- and memory-consuming because we have to check whether or not the copies of a Markov chain used for CFTP coalesce at a single state every time we extend the sample paths of the copies to the past.

Propp and Wilson (1996) stated that CFTP is effectively achieved by a monotone Markov chain (see, e.g., Keilson and Kester 1977) constructed from the target distribution, which is called *monotone CFTP* or *monotonic CFTP* (MCFTP). As far as we know, there have been a small number of examples for which MCFTP algorithms are established, for example, attractive spin systems (Propp and Wilson 1996), closed Jackson networks (Kijima and Matsui 2008a,b), discretized Dirichlet distributions (Matsui et al. 2010) and truncated Gaussian distributions (Philippe and Robert 2003). In particular, the algorithms proposed by Kijima and Matsui (2008a,b) and Matsui et al. (2010) are remarkably fast, though they are somewhat *sophisticated*.

*This paper has been submitted for publication in a special issue on “Queueing Theory and Network Applications” in Annals of Operations Research.

[†]E-mail: masuyama@sys.i.kyoto-u.ac.jp

The main purpose of this paper is to establish simple perfect samplers, which draw samples from an arbitrary target distribution $\{\pi(i); i \in \mathbb{S}\}$ on an arbitrary finite discrete set \mathbb{S} . It should be noted that \mathbb{S} is mapped one-to-one to a finite set of nonnegative numbers. Thus, we assume, without loss of generality, that $\mathbb{S} = \{0, 1, \dots, N\} =: \mathbb{Z}_N$, where N is a positive integer. We also assume that

$$\min_{i \in \mathbb{Z}_N} \pi(i) > 0. \quad (1.1)$$

For later use, let $\mathbb{N} = \{1, 2, 3, \dots\}$, $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ and $\mathbb{Z}_n = \{0, 1, \dots, n\}$ for any $n \in \mathbb{Z}_+$. For $n, m \in \mathbb{Z}$ such that $n \leq m$, let $\mathbb{Z}_{[n, m]} = \{n, n+1, \dots, m-1, m\}$. Let $x \vee y = \max(x, y)$ and $x \wedge y = \min(x, y)$ for $x, y \in (-\infty, \infty)$. Furthermore, we use the notation $f(x) = O(g(x))$ to represent $\limsup_{x \rightarrow \infty} |f(x)|/|g(x)| < \infty$.

In this paper, we first construct a monotone birth-and-death process (monotone BD-process or MBD for short) whose stationary distribution is equal to the target distribution $\{\pi(i); i \in \mathbb{Z}_N\}$. More specifically, we construct a monotone stochastic matrix $\mathbf{P} := (P(i, j))_{i, j \in \mathbb{Z}_N}$ such that

$$\mathbf{P} = \begin{pmatrix} r_0 & p_0 & 0 & \cdots & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 \\ 0 & q_2 & r_2 & p_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & q_{N-1} & r_{N-1} & p_{N-1} \\ 0 & \cdots & \cdots & 0 & q_N & r_N \end{pmatrix}, \quad (1.2)$$

where

$$p_i = \begin{cases} \frac{1}{1 + \gamma(0)}, & i = 0, \\ \frac{1}{1 + \gamma(i) \vee \gamma(i-1)}, & i \in \mathbb{Z}_{[1, N-1]}, \end{cases} \quad (1.3)$$

$$q_i = \begin{cases} \frac{\gamma(0)}{1 + \gamma(0)}, & i = 1, \\ \frac{\gamma(i-1)}{1 + \gamma(i-1) \vee \gamma(i-2)}, & i \in \mathbb{Z}_{[2, N]}, \end{cases} \quad (1.4)$$

$$\gamma(i) = \frac{\pi(i)}{\pi(i+1)}, \quad i \in \mathbb{Z}_{N-1}. \quad (1.5)$$

By definition, $r_i = 1 - p_i - q_i$ for $i \in \mathbb{Z}_N$ and $q_0 = p_N = 0$. We prove that \mathbf{P} is an irreducible and monotone stochastic matrix whose stationary distribution is equal to the target distribution $\{\pi(i); i \in \mathbb{Z}_N\}$ (see Theorem 2.1 below). We then discuss the first time when the copies of the MBD characterized by \mathbf{P} coalesce at a single state, which is called the *coalescence time* and denoted by T_C . Utilizing the existing results on BD-processes, we derive the upper bound for the expected coalescence time:

$$\mathbb{E}[T_C] \leq \theta N, \quad (1.6)$$

where $\theta \in (0, \infty)$ is a certain parameter (possibly depending on N).

Next we consider Doubling CFTP and Read-once CFTP (see, e.g., Huber 2016) using our MBD, which is referred to as *Doubling-MBD sampler* and *Read-once-MBD sampler*, respectively.

Using (1.6), we obtain upper bounds for the expected values and tail probabilities of the running times of Doubling-MBD and Read-once-MBD samplers. These upper bounds show that the expected running times of the two MBD samplers are $O(\theta N)$, and thus they are slower than the sophisticated special-purpose algorithms mentioned above. However, the construction of our MBD is very simple and little memory-consuming. In general, Doubling MCFTP and Read-once MCFTP are easily implementable (for details, see, e.g., Huber 2016). Therefore, Doubling-MBD and Read-once-MBD samplers are easily implementable and general-purpose perfect sampling algorithm. Furthermore, Read-once-MBD sampler is little memory-consuming, though the sampler is somewhat more time-consuming than Doubling-MBD sampler. As a result, Read-once-MBD sampler can draw samples from unnormalized target distributions with little memory consumption. This is a remarkable feature of Read-once-MBD sampler.

The rest of this paper is divided into two sections. Section 2 discusses our MBD constructed from the target distribution. Section 3 considers the performance of the two perfect samplers using our MBD.

2 Monotone BD-process from the target distribution

This section consists of two subsections. Section 2.1 constructs a monotone BD-process (MBD) whose stationary distribution is equal to the target distribution. Section 2.2 derives some upper bounds for the expected coalescence time of the copies of the MBD.

2.1 Construction of a monotone BD-process from the target distribution

The following theorem is the fundamental result of this paper.

Theorem 2.1 *The stochastic matrix \mathbf{P} defined by (1.2) together with (1.3)–(1.5) is an irreducible and monotone one whose stationary distribution is equal to the target distribution $\{\pi(i); i \in \mathbb{Z}_N\}$.*

Proof. From (1.1) and (1.3)–(1.5), we have

$$\begin{aligned} p_i &> 0, & i &\in \mathbb{Z}_{N-1}, \\ q_i &> 0, & i &\in \mathbb{Z}_{[1,N]}, \\ \pi(i)q_i &= \pi(i-1)p_{i-1}, & i &\in \mathbb{Z}_{[1,N]}, \end{aligned} \tag{2.1}$$

which show that \mathbf{P} is an irreducible stochastic matrix and that the target distribution $\{\pi(i); i \in \mathbb{Z}_N\}$ is a reversible measure and thus a unique stationary distribution of \mathbf{P} . Therefore, it suffices to prove that

$$\begin{aligned} p_i &\leq 1 - q_{i+1}, & i &\in \mathbb{Z}_{N-1}, \\ p_i &\leq 1 - q_i, & i &\in \mathbb{Z}_{N-1}, \end{aligned} \tag{2.2}$$

where (2.2) is the condition for the monotonicity of \mathbf{P} (see, e.g., Keilson and Kester 1977, Definition 1.2).

From (1.3) and (1.5), we have, for $i \in \mathbb{Z}_{N-1}$,

$$p_i \leq \frac{1}{1 + \gamma(i)} = \frac{\pi(i+1)}{\pi(i+1) + \pi(i)}, \quad (2.3)$$

and thus

$$\begin{aligned} p_i &\leq 1 - \frac{\pi(i)}{\pi(i+1) + \pi(i)} = 1 - \frac{\pi(i)}{\pi(i+1)} \frac{\pi(i+1)}{\pi(i+1) + \pi(i)} \\ &\leq 1 - \frac{\pi(i)}{\pi(i+1)} p_i = 1 - q_{i+1}, \end{aligned}$$

where the last inequality follows from (2.3) and the last equality follows from (2.1). Similarly, for $i \in \mathbb{Z}_{N-1}$,

$$\begin{aligned} p_i &\leq \frac{1}{1 + \gamma(i-1)} = \frac{\pi(i)}{\pi(i) + \pi(i-1)} \\ &= 1 - \frac{\pi(i-1)}{\pi(i) + \pi(i-1)} = 1 - \frac{\pi(i-1)}{\pi(i)} \frac{\pi(i)}{\pi(i) + \pi(i-1)} \\ &\leq 1 - \frac{\pi(i-1)}{\pi(i)} p_{i-1} = 1 - q_i. \end{aligned}$$

The proof is completed. \square

The following corollary is immediate from Theorem 2.1.

Corollary 2.1 *Suppose that the conditions of Theorem 2.1 are satisfied. We then have the following:*

(i) *If $\{\gamma(i); i \in \mathbb{Z}_{N-1}\}$ is nondecreasing, i.e.,*

$$\gamma(0) \leq \gamma(1) \leq \cdots \leq \gamma(N-1),$$

then (1.3) and (1.4) are reduced to

$$p_i = \frac{1}{1 + \gamma(i)}, \quad i \in \mathbb{Z}_{N-1}, \quad (2.4)$$

$$q_i = \frac{\gamma(i-1)}{1 + \gamma(i-1)}, \quad i \in \mathbb{Z}_{[1,N]}. \quad (2.5)$$

(ii) *If $\{\gamma(i); i \in \mathbb{Z}_{N-1}\}$ is nonincreasing, i.e.,*

$$\gamma(0) \geq \gamma(1) \geq \cdots \geq \gamma(N-1), \quad (2.6)$$

then (1.3) and (1.4) are reduced to

$$\begin{aligned} p_i &= \begin{cases} \frac{1}{1 + \gamma(0)}, & i = 0, \\ \frac{1}{1 + \gamma(i-1)}, & i \in \mathbb{Z}_{[1,N-1]}, \end{cases} \\ q_i &= \begin{cases} \frac{\gamma(0)}{1 + \gamma(0)}, & i = 1, \\ \frac{\gamma(i-1)}{1 + \gamma(i-1)}, & i \in \mathbb{Z}_{[2,N]}. \end{cases} \end{aligned}$$

Remark 2.1 Suppose that the conditions of the statement (i) of Corollary 2.1 are satisfied. Let $\{\hat{Y}_n; n \in \mathbb{Z}_+\}$ denote an MBD with state space \mathbb{Z}_N and transition probability matrix \mathbf{P} in (1.2) together with (2.4) and (2.5). Furthermore, suppose that the BD-processes $\{\hat{Y}_n\}$ starts with an initial distribution $\{\hat{\pi}(i); i \in \mathbb{Z}_N\}$ such that

$$\frac{\hat{\pi}(0)}{\pi(0)} \geq \frac{\hat{\pi}(1)}{\pi(1)} \geq \cdots \geq \frac{\hat{\pi}(N)}{\pi(N)}.$$

Note here that (2.6) yields

$$\pi(i-1)\pi(i+1) \leq [\pi(i)]^2, \quad i \in \mathbb{Z}_{[1, N-1]},$$

which shows that the target distribution $\{\pi(i)\}$ is log-concave. Therefore, it follows from Fill and Kahn (2013, Proposition 3.2, Corollary 3.3(a) and Theorem 5.1) that the BD-process $\{\hat{Y}_n\}$ mixes (i.e., converges to stationarity) faster in total variation distance than does an arbitrary MBD $\{\hat{Z}_n; n \in \mathbb{Z}_+\}$ that has the same state space \mathbb{Z}_N , stationary distribution $\{\pi(i); i \in \mathbb{Z}_N\}$ and initial distribution $\{\hat{\pi}(i); i \in \mathbb{Z}_N\}$ as the BD-process $\{\hat{Y}_n; n \in \mathbb{Z}_+\}$.

Next we describe a construction of the copies of the MBD with state space \mathbb{Z}_N and transition probability matrix \mathbf{P} , which can be used for MCFTP. To this end, we define $\{U_m; m \in \mathbb{Z}\}$ as a sequence of independent and identically distributed (i.i.d.) uniform random variables in $(0, 1)$. We then have the following result.

Theorem 2.2 Suppose that the conditions of Theorem 2.1 are satisfied. Let $\phi : \mathbb{Z}_N \times (0, 1) \rightarrow \mathbb{Z}_N$ denote a function such that, for $i \in \mathbb{Z}_N$ and $u \in (0, 1)$,

$$\phi(i, u) = \begin{cases} i+1, & u \in (1-p_i, 1), \\ i, & u \in [q_i, 1-p_i], \\ i-1, & u \in (0, q_i), \end{cases} \quad (2.7)$$

where p_i and q_i are given in (1.3) and (1.4), respectively. Furthermore, for each $k \in \mathbb{Z}_N$, let $\{X_n^{(k)}; n \in \mathbb{Z}_+\}$ denote a sequence of random variables such that

$$X_n^{(k)} = \begin{cases} k, & n = 0, \\ \phi(X_{n-1}^{(k)}, U_n), & n \in \mathbb{N}. \end{cases} \quad (2.8)$$

Under these conditions, the stochastic processes $\{X_n^{(k)}; n \in \mathbb{Z}_+\}$'s, $k \in \mathbb{Z}_N$, are MBDs with transition probability matrix \mathbf{P} , which satisfy

$$X_n^{(0)} \leq X_n^{(1)} \leq \cdots \leq X_n^{(N)} \quad \text{for all } n \in \mathbb{N}. \quad (2.9)$$

Proof. It is clear that $\{X_n^{(k)}; n \in \mathbb{Z}_+\}$'s, $k \in \mathbb{Z}_N$, are MBDs with transition probability matrix \mathbf{P} . Thus, we prove that (2.9) holds.

It follows from (2.7) that, for $i \in \mathbb{Z}_{N-1}$,

$$\begin{aligned} \phi(i+1, u) &\geq i+1 \geq \phi(i, u), & 1 > u \geq q_{i+1}, \\ \phi(i+1, u) &\geq i \geq \phi(i, u), & 0 < u \leq 1-p_i. \end{aligned}$$

It also follows from (2.2) that $q_{i+1} \leq 1 - p_i$ for $i \in \mathbb{Z}_{N-1}$. Therefore,

$$\phi(i+1, u) \geq \phi(i, u), \quad i \in \mathbb{Z}_{N-1}, u \in (0, 1). \quad (2.10)$$

Combining (2.8) and (2.10) yields (2.9). \square

Theorem 2.2 shows that the function ϕ , together with the uniform random variables U_m 's, generates MBDs with transition probability matrix \mathbf{P} . Thus, we refer to ϕ as a *monotone update function* for MBDs with \mathbf{P} . Note here that $\{X_n^{(k)}; n \in \mathbb{Z}_+\}$'s can be considered the copies of a generic BD-process driven by the monotone update function ϕ , which is denoted by $\{X_n; n \in \mathbb{Z}_+\}$. Especially, we refer to $\{X_n^{(N)}\}$ and $\{X_n^{(0)}\}$ as the *upper-bounding* and *lower-bounding* copies, respectively, of $\{X_n\}$.

2.2 Expected coalescence time of the copies of the monotone BD-process

Let T_C denote

$$T_C = \inf\{n \in \mathbb{N} : X_n^{(0)} = X_n^{(1)} = \cdots = X_n^{(N)}\}, \quad (2.11)$$

which is the first time when all the copies $\{X_n^{(k)}\}$'s coalesce at a single state in the state space \mathbb{Z}_N . Thus, we call T_C the *coalescence time* of the copies $\{X_n^{(k)}\}$'s of $\{X_n\}$. It follows from (2.9) and (2.11) that

$$T_C = \inf\{n \in \mathbb{N} : X_n^{(0)} = X_n^{(N)}\}. \quad (2.12)$$

We now define $T_{i,j}$, $i, j \in \mathbb{Z}_N^2$, $i \neq j$, as a generic random variable for the first passage time from state i to state j . We assume that $\{T_{0,1}, T_{1,2}, \dots, T_{N-1,N}\}$ are independent and so are $\{T_{N,N-1}, T_{N-1,N-2}, \dots, T_{1,0}\}$, which does not lose generality due to the skip-free property of BD-processes. It then follows from (2.12) that

$$\begin{aligned} T_C &\leq \inf\{n \in \mathbb{N} : X_n^{(0)} = N\} \stackrel{d}{=} T_{0,N} \stackrel{d}{=} \sum_{i=0}^{N-1} T_{i,i+1}, \\ T_C &\leq \inf\{n \in \mathbb{N} : X_n^{(N)} = 0\} \stackrel{d}{=} T_{N,0} \stackrel{d}{=} \sum_{i=0}^{N-1} T_{i+1,i}, \end{aligned}$$

where the symbol " $\stackrel{d}{=}$ " represents the equality in distribution. Therefore,

$$\mathbb{E}[T_C] \leq \mathbb{E}[T_{0,N}] \wedge \mathbb{E}[T_{N,0}]. \quad (2.13)$$

We can readily obtain (see, e.g., Theorem 4.11 of Heyman and Sobel (2004), where continuous-time BD-processes are considered)

$$\mathbb{E}[T_{0,N}] = \sum_{i=0}^{N-1} \mathbb{E}[T_{i,i+1}] = \sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=0}^i \frac{\pi(m)}{\pi(i)}. \quad (2.14)$$

$$\mathbb{E}[T_{N,0}] = \sum_{i=0}^{N-1} \mathbb{E}[T_{i+1,i}] = \sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=i+1}^N \frac{\pi(m)}{\pi(i)}. \quad (2.15)$$

Substituting (2.14) and (2.15) into (2.13) yields

$$\mathbb{E}[T_C] \leq \left(\sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=0}^i \frac{\pi(m)}{\pi(i)} \right) \wedge \left(\sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=i+1}^N \frac{\pi(m)}{\pi(i)} \right). \quad (2.16)$$

Using (2.16), we obtain the following result.

Theorem 2.3 *If the conditions of Theorem 2.1 are satisfied, then*

$$\mathbb{E}[T_C] \leq \theta N, \quad (2.17)$$

where θ is a positive constant such that

$$\theta = \left[\max_{i \in \mathbb{Z}_{N-1}} \left(\frac{1}{p_i} \sum_{m=0}^i \frac{\pi(m)}{\pi(i)} \right) \right] \wedge \left[\max_{i \in \mathbb{Z}_{N-1}} \left(\frac{1}{p_i} \sum_{m=i+1}^N \frac{\pi(m)}{\pi(i)} \right) \right]. \quad (2.18)$$

Proof. Note that

$$\begin{aligned} \sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=0}^i \frac{\pi(m)}{\pi(i)} &\leq \max_{i \in \mathbb{Z}_{N-1}} \left(\frac{1}{p_i} \sum_{m=0}^i \frac{\pi(m)}{\pi(i)} \right) N, \\ \sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=i+1}^N \frac{\pi(m)}{\pi(i)} &\leq \max_{i \in \mathbb{Z}_{N-1}} \left(\frac{1}{p_i} \sum_{m=i+1}^N \frac{\pi(m)}{\pi(i)} \right) N. \end{aligned}$$

Substituting these inequalities into (2.16) leads to (2.17) with (2.18). \square

Under some additional conditions, we obtain simpler bounds for $\mathbb{E}[T_C]$.

Theorem 2.4 *Suppose that the conditions of Theorem 2.1 are satisfied. We then have the following:*

(i) *If there exists some $C \in (0, \infty)$ independent of N such that*

$$\max_{i \in \mathbb{Z}_{N-1}} \sum_{m=0}^i \frac{\pi(m)}{\pi(i)} \leq C \quad \text{or} \quad \max_{i \in \mathbb{Z}_{N-1}} \sum_{m=i+1}^N \frac{\pi(m)}{\pi(i)} \leq C, \quad (2.19)$$

then

$$\mathbb{E}[T_C] \leq C \left(\max_{i \in \mathbb{Z}_{N-1}} \frac{1}{p_i} \right) N. \quad (2.20)$$

(ii) *If there exists some $C \in (0, \infty)$ independent of N such that*

$$\max_{i \in \mathbb{Z}_{N-1}} \max_{0 \leq m \leq i} \frac{\pi(m)}{\pi(i)} \leq C \quad \text{or} \quad \max_{i \in \mathbb{Z}_{N-1}} \max_{i+1 \leq m \leq N} \frac{\pi(m)}{\pi(i)} \leq C, \quad (2.21)$$

then

$$\mathbb{E}[T_C] \leq C \left(\max_{i \in \mathbb{Z}_{N-1}} \frac{1}{p_i} \right) \frac{N(N+1)}{2}. \quad (2.22)$$

Remark 2.2 If $\{\pi(i); i \in \mathbb{Z}_N\}$ is nonincreasing or nondecreasing, then (2.21) holds for $C = 1$ and thus the statement (ii) of Theorem 2.4 yields

$$\mathbb{E}[T_C] \leq \left(\max_{i \in \mathbb{Z}_{N-1}} \frac{1}{p_i} \right) \frac{N(N+1)}{2}. \quad (2.23)$$

Proof of Theorem 2.4. We first prove the statement (i). Applying (2.19) to (2.16) yields

$$\mathbb{E}[T_C] \leq C \sum_{i=0}^{N-1} \frac{1}{p_i} \leq C \left(\max_{i \in \mathbb{Z}_{N-1}} \frac{1}{p_i} \right) N,$$

which shows that (2.20) holds. Next we prove the statement (ii). Combining (2.21) and (2.16), we have either of the following inequalities:

$$\mathbb{E}[T_C] \leq \sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=0}^i C \leq C \left(\max_{i \in \mathbb{Z}_{N-1}} \frac{1}{p_i} \right) \sum_{i=0}^{N-1} (i+1),$$

$$\mathbb{E}[T_C] \leq \sum_{i=0}^{N-1} \frac{1}{p_i} \sum_{m=i+1}^N C \leq C \left(\max_{i \in \mathbb{Z}_{N-1}} \frac{1}{p_i} \right) \sum_{i=0}^{N-1} (N-i).$$

Each of the two inequalities shows that (2.22) holds. \square

Example 2.1 (Truncated geometric distribution) Consider a truncated geometric distribution. To this end, fix

$$\pi(i) = \frac{(1-\xi)\xi^i}{1-\xi^{N+1}}, \quad i \in \mathbb{Z}_N,$$

where $0 < \xi < 1$. Clearly, $\gamma(i) = \xi^{-1}$ for $i \in \mathbb{Z}_{N-1}$, which satisfies the conditions of the statement (i) of Corollary 2.1. Thus, from (2.4) and (2.5), we have

$$p_i = \frac{1}{1 + \gamma(i)} = \frac{\xi}{1 + \xi}, \quad i \in \mathbb{Z}_{N-1}.$$

Note here that

$$\begin{aligned} \max_{i \in \mathbb{Z}_{N-1}} \sum_{m=0}^i \frac{\pi(m)}{\pi(i)} &= \max_{i \in \mathbb{Z}_{N-1}} \sum_{m=0}^i \xi^{m-i} \leq \frac{1}{1-\xi}, \\ \max_{i \in \mathbb{Z}_{N-1}} \sum_{m=i+1}^N \frac{\pi(m)}{\pi(i)} &= \max_{i \in \mathbb{Z}_{N-1}} \sum_{m=i+1}^N \xi^{m-i} \leq \frac{\xi}{1-\xi}. \end{aligned}$$

Combining these results and the statement (i) of Theorem 2.4 yields

$$\mathbb{E}[T_C] \leq \frac{\xi}{1-\xi} \frac{1+\xi}{\xi} N = \frac{1+\xi}{1-\xi} N.$$

Example 2.2 (Zipf distribution) Consider the following Zipf distribution $\{\pi(i); i \in \mathbb{Z}_N\}$:

$$\pi(i) = \frac{(i+1)^{-\alpha}}{\sum_{\ell=0}^N (\ell+1)^{-\alpha}}, \quad i \in \mathbb{Z}_N,$$

where $\alpha > 1$. We then have

$$\gamma(i) = \left(1 + \frac{1}{i+1}\right)^\alpha, \quad i \in \mathbb{Z}_{N-1},$$

which is decreasing with i . Therefore, according to the statement (ii) of Corollary 2.1, we fix

$$p_i = \frac{1}{1 + \gamma((i-1) \vee 0)} = \frac{(i \vee 1)^\alpha}{(i \vee 1)^\alpha + \{(i \vee 1) + 1\}^\alpha}, \quad i \in \mathbb{Z}_{N-1}. \quad (2.24)$$

Furthermore, since $\{\pi(i); i \in \mathbb{Z}_N\}$ is decreasing (see Remark 2.2), it follows from (2.23) and (2.24) that

$$\mathbb{E}[T_C] \leq \max_{i \in \mathbb{Z}_{[1, N-1]}} \left[\frac{i^\alpha + (i+1)^\alpha}{i^\alpha} \right] \frac{N(N+1)}{2} \leq (1 + 2^\alpha) \frac{N(N+1)}{2}.$$

3 Perfect samplers using the monotone BD-process

In this section, we discuss the running times of Doubling CFTP and Read-once CFTP using the monotone update function ϕ , which are referred to as *Doubling-MBD sampler* and *Read-once-MBD sampler*, respectively.

To facilitate the subsequent discussion, we introduce some definitions. For $m \in \mathbb{Z}$ and $n \in \mathbb{Z}_+$, let

$$\mathbf{U}_m^{(n)} = (U_m, U_{m+1}, \dots, U_{m+n-1}).$$

For convenience, let $\mathbf{U}_m^{(-n)} = \emptyset$ for all $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. In addition, for $s \in \mathbb{Z}$, $n \in \mathbb{Z}_+$ and $x \in \mathbb{Z}_N$, let $\Phi_s^{s+n}(x, \mathbf{U}_m^{(n)})$ denote

$$\Phi_s^{s+n}(x, \mathbf{U}_m^{(n)}) = \phi(\phi(\dots \phi(x, U_m), \dots, U_{m+n-2}), U_{m+n-1}),$$

where ϕ is the monotone update function given in (2.7) and $\{U_m; m \in \mathbb{Z}\}$ is a sequence of i.i.d. uniform random variables in $(0, 1)$. Note that, for any $t \in \mathbb{Z}_+$, the two processes $\{\Phi_{-t}^{-t+n}(N, \mathbf{U}_{-t}^{(n)}); n \in \mathbb{Z}_+\}$ and $\{\Phi_{-t}^{-t+n}(0, \mathbf{U}_{-t}^{(n)}); n \in \mathbb{Z}_+\}$ are the upper- and lower-bounding copies of an MBD with transition probability matrix \mathbf{P} , which run from time $-t$ to time $-t + n$.

We first consider Doubling MBD sampler, which is described in Algorithm 1.

Algorithm 1: Doubling-MBD sampler

Output: X

(i) Set $t = 2$.

(ii) Double t until

$$Y := \Phi_{-t}^{-t/2}(0, \mathbf{U}_{-t}^{(t/2)}) = \Phi_{-t}^{-t/2}(N, \mathbf{U}_{-t}^{(t/2)}).$$

(iii) Return $X = \Phi_{-t/2}^0(Y, \mathbf{U}_{-t/2}^{(t/2)})$.

Let T_D denote the number of the uniform random variables used by Algorithm 1, i.e., T_D is equal to a positive integers such that

$$T_D = \inf\{t \in \mathbb{N} : \Phi_{-t}^{-t/2}(0, \mathbf{U}_{-t}^{(t/2)}) = \Phi_{-t}^{-t/2}(N, \mathbf{U}_{-t}^{(t/2)})\}.$$

Following Huber (2008), we read T_D as the running time of Algorithm 1. Using Huber (2008, Lemma 5.4), we obtain the following result.

Proposition 3.1 (Doubling-MBD sampler)

$$\mathbb{E}[T_D] \leq 4\theta N, \quad (3.1)$$

$$\mathbb{P}(T_D > k\theta N) \leq \exp\{1 - k/(4e)\}, \quad k \in \mathbb{Z}_+, \quad (3.2)$$

where θ is the positive constant given in (2.18).

Proof. It follows from Huber (2008, Lemma 5.4) that

$$\mathbb{E}[T_D] \leq 4\mathbb{E}[T_C],$$

$$\mathbb{P}(T_D > k\mathbb{E}[T_C]) \leq \exp\{1 - k/(4e)\}, \quad k \in \mathbb{Z}_+.$$

Combining these with Theorem 2.3 results in (3.1) and (3.2). \square

Next we consider Read-once-MBD sampler, which is described in Algorithm 2 below.

Algorithm 2: Read-once-MBD sampler

Input: Block size $B \in \mathbb{N}$

Output: X

(i) Set $\ell = 1$.

(ii) If

$$\Phi_0^B(0, \mathbf{U}_{(\ell-1)B}^{(B)}) \neq \Phi_0^B(N, \mathbf{U}_{(\ell-1)B}^{(B)}),$$

then increment ℓ by one and go back to Step (ii); otherwise set $X = \Phi_0^B(0, \mathbf{U}_{(\ell-1)B}^{(B)})$ and go to Step (iii) with $\ell' = 1$.

(iii) Set $Y = X$ and perform the following: If

$$\Phi_{B+(\ell'-1)B}^{B+\ell'B}(0, \mathbf{U}_{\ell B+(\ell'-1)B}^{(B)}) \neq \Phi_{B+(\ell'-1)B}^{B+\ell'B}(N, \mathbf{U}_{\ell B+(\ell'-1)B}^{(B)}),$$

then set $X = \Phi_{B+(\ell'-1)B}^{B+\ell'B}(Y, \mathbf{U}_{\ell B+(\ell'-1)B}^{(B)})$ and go back to Step (iii) with incrementing ℓ' by one; otherwise return X .

Remark 3.1 When Algorithm 2 stops, we have

$$X = \Phi_0^{\ell'B}(0, \mathbf{U}_{(\ell-1)B}^{\ell'B}) = \Phi_0^{\ell'B}(N, \mathbf{U}_{(\ell-1)B}^{\ell'B}).$$

As with Algorithm 1, we define T_R as the number of the uniform random variables used by Algorithm 2, and then read T_R as the running time of Algorithm 2. Let L and L' denote the

numbers of the iterations in Steps (ii) and (iii), respectively, of Algorithm 2. By definition, L and L' are independent and

$$P(L > k) = P(L' > k) = [P(T_C > B)]^k, \quad k \in \mathbb{Z}_+. \quad (3.3)$$

In addition,

$$T_R = (L + L')B. \quad (3.4)$$

Using Theorem 2.3 together with (3.3) and (3.4), and proceeding as in the proof of Huber (2008, Lemma 5.4), we obtain the following result.

Theorem 3.1 (Read-once-MBD sampler) *Fix $b \in \mathbb{N}$ such that $b > e$, and fix the block size B of Algorithm 2 such that $B = b\lceil\theta\rceil N$, where θ is the positive constant given in (2.18). We then have*

$$E[T_R] \leq \frac{2b\lceil\theta\rceil N}{1 - \beta(b)}, \quad (3.5)$$

$$P(T_R > b\lceil\theta\rceil Nk) \leq (1 - \beta(b))[\beta(b)]^{k-1}k + [\beta(b)]^k, \quad k \in \mathbb{N}, \quad (3.6)$$

where $\beta(b) = \exp\{1 - b/e\} \in (0, 1)$. In addition, the value of integer $b > e$ minimizing the right hand side of (3.5) is equal to six, or equivalently,

$$\arg \min_{b \in \{3, 4, 5, \dots\}} \frac{2b\lceil\theta\rceil N}{1 - \beta(b)} = 6. \quad (3.7)$$

Proof. It follows from Markov's inequality that, for any fixed $\alpha > 1$,

$$P(T_C > \alpha E[T_C]) \leq \alpha^{-1}. \quad (3.8)$$

Note here that $\{P(T_C > x); x \geq 0\}$ is log-subadditive (see, e.g., Propp and Wilson 1996, Theorem 6). Thus, from (3.8), we have

$$\begin{aligned} P(T_C > kE[T_C]) &\leq \left(\frac{1}{\alpha}\right)^{\lfloor k/\alpha \rfloor} \leq \alpha \left(\frac{1}{\alpha}\right)^{k/\alpha} \\ &= \alpha \exp\{-k(\ln \alpha)/\alpha\}, \quad k \in \mathbb{Z}_+. \end{aligned} \quad (3.9)$$

We now fix $\alpha = e$ to maximizing $(\ln \alpha)/\alpha$. It then follows from (3.9) that

$$P(T_C > kE[T_C]) \leq \exp\{1 - k/e\} = \beta(k), \quad k \in \mathbb{Z}_+. \quad (3.10)$$

From $B = b\lceil\theta\rceil N$ and Theorem 2.3, we also have $B \geq bE[T_C]$. Using this and (3.10), we obtain

$$P(T_C > B) \leq P(T_C > bE[T_C]) \leq \beta(b). \quad (3.11)$$

Substituting (3.11) into (3.3) yields

$$P(L > k) = P(L' > k) \leq [\beta(b)]^k, \quad k \in \mathbb{Z}_+.$$

Therefore, there exist independent random variables \bar{L} and \bar{L}' such that

$$L \leq \bar{L}, \quad L' \leq \bar{L}', \quad (3.12)$$

$$P(\bar{L} > k) = P(\bar{L}' > k) = [\beta(b)]^k, \quad k \in \mathbb{Z}_+. \quad (3.13)$$

It follows from (3.13) that

$$\mathbb{E}[L] + \mathbb{E}[L'] = \frac{2}{1 - \beta(b)},$$

and

$$\begin{aligned} \mathbb{P}(\bar{L} + \bar{L}' > k) &= \sum_{m=1}^k \mathbb{P}(\bar{L} = m) \mathbb{P}(\bar{L}' > k - m) + \mathbb{P}(\bar{L} > k) \\ &= (1 - \beta(b)) \sum_{m=1}^k [\beta(b)]^{m-1} [\beta(b)]^{k-m} + [\beta(b)]^k \\ &= (1 - \beta(b)) [\beta(b)]^{k-1} k + [\beta(b)]^k, \quad k \in \mathbb{N}. \end{aligned}$$

Combining these results with (3.4) and (3.12), we have

$$\begin{aligned} \mathbb{E}[T_R] &\leq B \cdot \mathbb{E}[\bar{L} + \bar{L}'] = \frac{2B}{1 - \beta(b)}, \\ \mathbb{P}(T_R > kB) &\leq \mathbb{P}(\bar{L} + \bar{L}' > k) = (1 - \beta(b)) [\beta(b)]^{k-1} k + [\beta(b)]^k, \quad k \in \mathbb{N}, \end{aligned}$$

which imply that (3.5) and (3.6) hold due to $B = b[\theta]N$.

In what follows, we prove (3.7), which is equivalent to

$$\arg \min_{x \in \{3, 4, 5, \dots\}} F(x) = 6, \quad (3.14)$$

where F denotes a function such on (e, ∞) that

$$F(x) = \frac{x}{1 - \beta(x)} = \frac{x}{1 - \exp\{1 - x/e\}}, \quad x > e.$$

By definition, F is convex and

$$F'(x) = \frac{1 - \exp\{1 - x/e\} - e^{-1}x \exp\{1 - x/e\}}{[1 - \exp\{1 - x/e\}]^2}, \quad x > e.$$

Let $G(x)$, $x > e$, denote the numerator of $F'(x)$ in the above equation, i.e.,

$$G(x) = 1 - \exp\{1 - x/e\} - e^{-1}x \exp\{1 - x/e\}, \quad x > e.$$

We then have

$$\begin{aligned} G(2e) &= 1 - 3e^{-1} < 0, \\ G(2.5e) &= 1 - 3.5(e\sqrt{e})^{-1} > 1 - 3.5 \times (2.5 \times 1.5)^{-1} = \frac{1}{15} > 0, \end{aligned}$$

which lead to $F'(2e) < 0$ and $F'(2.5e) > 0$. Note here that $2e > 5.4$ and $2.5e < 7$. Therefore, the convexity of F yields $F'(5) < 0$ and $F'(7) > 0$, which results in (3.14). \square

Proposition 3.1 and Theorem 3.1 imply that the running time T_D of Doubling-MBD sampler is less than the running time T_R of Read-once-MBD sampler. However, Doubling-MBD sampler has to store all the generated (uniform) random numbers until it outputs a sample following the target

distribution. On the other hand, Read-once-MBD sampler is little memory-consuming because the sampler uses, only one time, each of the generated random numbers.

We close this section by comparing our perfect samplers with the inverse transform sampling (see, e.g., Fishman 1996). The inverse transform sampling for discrete target distributions is easy implementable and takes the $O(N)$ running time in order to draw a sample from the target distribution. Therefore, the inverse transform sampling is less time-consuming than our perfect samplers.

To discuss this topic from a different perspective, we suppose that the target distribution $\{\pi(i); i \in \mathbb{Z}_N\}$ is not normalized, in other words, we have an unnormalized target distribution $\{\hat{\pi}(i); i \in \mathbb{Z}_N\}$ such that $C_\pi := \sum_{i=0}^N \hat{\pi}(i) \neq 1$ and

$$\pi(i) = \frac{1}{C_\pi} \hat{\pi}(i), \quad i \in \mathbb{Z}_N. \quad (3.15)$$

It then follows from (1.5) and (3.15) that

$$\gamma(i) = \frac{\hat{\pi}(i)}{\hat{\pi}(i+1)}, \quad i \in \mathbb{Z}_{N-1}.$$

Therefore, our two perfect samplers still work well by using the unnormalized target distribution $\{\hat{\pi}(i)\}$. On the other hand, the inverse transform sampling has a problem in the present situation because it needs the cumulative distribution $\{\sigma(i); i \in \mathbb{Z}_N\}$, where $\sigma(i) = \sum_{\ell=0}^i \pi(\ell)$ for $i \in \mathbb{Z}_N$. To obtain the cumulative distribution $\{\sigma(i)\}$, we have to compute the normalizing constant C_π by summing the unnormalized target distribution $\{\hat{\pi}(i)\}$ over its support set \mathbb{Z}_N .

It should be note that the obtained constant C_π includes, at worst, the $O(N)$ rounding error. Such rounding error can be reduced to $O(\ln N)$ if C_π is computed by pairwise summation (see, e.g., Higham 1993). Furthermore, if C_π is computed by Kahan summation algorithm, then the rounding error can be basically reduced to $O(1)$ but its computational complexity is four times as much as that of *naive* summation (see, e.g., Higham 1993). Even though we take any of these options, we have to store all the information of the cumulative distribution $\{\sigma(i)\}$. Such memory consumption is not necessary for our two perfect samplers.

As a result, although our MBD perfect samplers may not be particularly superior in speed to other methods, they are easily implementable and can draw samples *exactly* from unnormalized target distributions. Especially, Read-one MBD sampler achieves such *exact* sampling with little memory consumption.

Acknowledgments

The author thanks Dr. Shuji Kijima for helpful comments that motivated this work.

References

- Fill, J. A., Kahn, J., 2013. Comparison inequalities and fastest-mixing Markov chains. The Annals of Applied Probability 23 (5), 1778–1816.
- Fishman, G., 1996. Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York.

- Heyman, D. P., Sobel, M. J., 2004. *Stochastic Models in Operations Research, Vol. I: Stochastic Processes and Operating Characteristics*, paperback Edition. Dover Publications, Mineola, New York.
- Higham, N. J., 1993. The accuracy of floating point summation. *SIAM Journal on Scientific Computing* 14 (4), 783–799.
- Huber, M., 2008. Perfect simulation with exponential tails on the running time. *Random Structures and Algorithms* 33 (1), 29–43.
- Huber, M., 2016. *Perfect Simulation*. CRC Press, Boca Raton, FL.
- Keilson, J., Kester, A., 1977. Monotone matrices and monotone Markov processes. *Stochastic Processes and their Applications* 5, 231–241.
- Kijima, S., Matsui, T., 2008a. Approximation algorithm and perfect sampler for closed jackson networks with single servers. *SIAM Journal on Computing* 38 (4), 1484–1503.
- Kijima, S., Matsui, T., 2008b. Randomized approximation scheme and perfect sampler for closed jackson networks with multiple servers. *Annals of Operations Research* 162 (1), 35–55.
- Matsui, T., Motoki, M., Kamatani, N., Kijima, S., 2010. Polynomial time approximate or perfect samplers for discretized Dirichlet distribution. *Japan Journal of Industrial and Applied Mathematics* 27 (1), 91–123.
- Philippe, A., Robert, C. P., 2003. Perfect simulation of positive Gaussian distributions. *Statistics and Computing* 13 (2), 179–186.
- Propp, J. G., Wilson, D. B., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random structures and Algorithms* 9 (1–2), 223–252.